

PCT

WELTORGANISATION FÜR GEISTIGES EIGENTUM  
Internationales Büro



INTERNATIONALE ANMELDUNG VERÖFFENTLICH NACH DEM VERTRAG ÜBER DIE  
INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

(51) Internationale Patentklassifikation <sup>7</sup> :  C12Q 1/68		A2	(11) Internationale Veröffentlichungsnummer: WO 00/44934  (43) Internationales Veröffentlichungsdatum: 3. August 2000 (03.08.00)
(21) Internationales Aktenzeichen: PCT/DE00/00288  (22) Internationales Anmeldedatum: 27. Januar 2000 (27.01.00)  (30) Prioritätsdaten: 199 05 082.1 29. Januar 1999 (29.01.99) DE		(81) Bestimmungsstaaten: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO Patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).  Veröffentlicht <i>Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.</i>	
(71) Anmelder ( <i>für alle Bestimmungsstaaten ausser US</i> ): EPIGENOMICS GMBH [DE/DE]; Kastanienallee 24, D-10435 Berlin (DE).  (72) Erfinder; und (75) Erfinder/Anmelder ( <i>nur für US</i> ): OLEK, Alexander [DE/DE]; Kyffhäuserstrasse 20, D-10781 Berlin (DE).  (74) Anwalt: SCHUBERT, Clemens; Joachimstrasse 9, D-10119 Berlin-Mitte (DE).			
(54) Title: METHOD OF IDENTIFYING CYTOSINE METHYLATION PATTERNS IN GENOMIC DNA SAMPLES			
(54) Bezeichnung: VERFAHREN ZUR IDENTIFIKATION VON CYTOSIN-METHYLIERUNGSMUSTERN IN GENOMISCHEN DNA-PROBEN			
(57) Abstract			
The invention relates to a method of identifying cytosine methylation patterns in genomic DNA samples. The inventive method comprises the following steps: a) chemically treating a genomic DNA sample in such a manner that the cytosine and the 5-methyl cytosine react differently and that the two products have a different base-pairing behavior in the duplex; b) enzymatically amplifying parts of the DNA sample treated in this manner; c) binding the amplified parts of the DNA sample treated in this manner to a surface; d) hybridizing a set of probes of different nucleic acid sequences to the immobilized DNA samples, said respective base sequences containing at least once the dinucleotide sequence 5'-CpG-3'; e) removing the non-hybridized probes; f) analyzing the hybridized probes in a mass spectrometer, the position of the probes on the sample carrier allowing an allocation of the DNA sample to be hybridized; g) converting the peak-patterns obtained from the mass-spectra to methylation patterns and matching the new data with a data library.			
(57) Zusammenfassung			
Die Erfindung betrifft ein Verfahren zur Identifikation von Cytosin-Methylierungsmustern in genomischen DNA-Proben, wobei man a) eine genetische DNA-Probe chemisch so behandelt, daß Cytosin und 5-Methylcytosin unterschiedlich reagieren und man in der Duplex ein unterschiedliches Basenpaarungsverhalten der beiden Produkte erhält; b) Teile der so behandelten DNA-Probe enzymatisch amplifiziert; c) die amplifizierten Teile der so behandelten DNA-Probe an eine Oberfläche bindet; d) einen Satz von Sonden unterschiedlicher Nukleotidsequenzen, die jeweils mindestens einmal die Dinukleotidsequenz 5'-CpG-3' enthalten, an die immobilisierten DNA-Proben hybridisiert; e) die nicht hybridisierten Sonden abtrennt; f) die hybridisierten Sonden in einem Massenspektrometer analysiert, wobei die Position der Sonden auf dem Probenträger eine Zuordnung zu der hybridisierenden DNA-Probe erlaubt; g) Übertragung der aus den Massenspektren erhaltenen Peakmuster in Methylierungsmuster und Abgleich der neuen Daten mit einer Datenbank.			

#### ***LEDIGLICH ZUR INFORMATION***

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidschan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		
EE	Estland						

**Verfahren zur Identifikation von Cytosin-Methylierungsmustern in genomischen DNA-Proben**

5 Die Erfindung betrifft ein Verfahren zur Identifikation von Cytosin-Methylierungsmustern in genomischen DNA-Proben.

10 Die genetische Information, die durch vollständige Sequenzierung genomischer DNA als Basenabfolge erhalten wird, beschreibt das Genom einer Zelle nur unvollständig. 5-Methylcytosin-Nukleobasen, die durch reversible Methylierung von DNA in der Zelle entstehen, sind ein epigenetischer Informationsträger und dienen beispielsweise zur 15 Regulation von Promotoren. Der Methylierungszustand eines Genoms repräsentiert den gegenwärtigen Status der Genexpression, ähnlich wie ein mRNA Expressionsmuster.

20 5-Methylcytosin ist die häufigste kovalent modifizierte Base in der DNA eukaryotischer Zellen. Sie spielt beispielsweise eine Rolle in der Regulation der Transkription, genomischem Imprinting und in der Tumorgenese. Die Identifizierung von 5-Methylcytosin als Bestandteil genetischer Information ist daher von erheblichem Interesse. 25 5-Methylcytosin-Positionen können jedoch nicht durch Sequenzierung identifiziert werden, da 5-Methylcytosin das gleiche Basenpaarungsverhalten aufweist wie Cytosin. Darüberhinaus geht bei einer PCR-Amplifikation die epigenetische Information, die die 5-Methylcytosine tragen, 30 vollständig verloren.

35 Es sind mehrere Verfahren bekannt, die diese Probleme zu lösen versuchen. Meist wird eine chemische Reaktion oder enzymatische Behandlung der genomischen DNA durchgeführt, in deren Folge sich die Cytosin- von den Methylcytosin-Basen unterscheiden lassen. Eine gängige Methode ist die

Umsetzung von genomischer DNA mit Bisulfit, die nach alkalischer Hydrolyse in zwei Schritten zu einer Umwandlung der Cytosin Basen in Uracil führt (Shapiro, R., Cohen, B., Servis, R. Nature 227, 1047 (1970). 5-Methylcytosin bleibt unter diesen Bedingungen unverändert. Die Umwandlung von C in U führt zu einer Veränderung der Basensequenz, aus der sich durch Sequenzierung nun die ursprünglichen 5-Methylcytosine ermitteln lassen (nur diese liefern noch eine Bande in der C-Spur).

10

Eine Übersicht über weitere bekannte Möglichkeiten, 5-Methylcytosine nachzuweisen, ist beispielsweise dem folgenden Übersichtsartikel zu entnehmen: Rein, T., DePamphilis, M. L., Zorbas, H., Nucleic Acids Res. 26, 2255 (1998).

15

Die Bisulphit-Technik wird bisher bis auf wenige Ausnahmen (z.B. Zeschnick, M. et al., Eur. J. Hum. Gen. 5, 94-98; Kubota T. et al., Nat. Genet. 16, 16-17) nur in der Forschung angewendet. Immer aber werden kurze, spezifische Stücke eines bekannten Gens nach einer Bisulphit-Behandlung amplifiziert und entweder komplett sequenziert (Olek, A. and Walter, J., Nat. Genet. 17, 275-276) oder einzelne Cytosin-Positionen durch eine „Primer-Extension-Reaktion“ (Gonzalgo, M. L. and Jones, P.A., Nucl. Acids. Res. 25, 2529-2531) oder Enzymschnitt (Xiong, Z. and Laird, P.W., Nucl. Acids. Res. 25, 2532-2534) nachgewiesen. Alle diese Referenzen stammen aus dem Jahre 1997. Das Konzept, komplexe Methylierungsmuster zur Korrelation mit phänotypischen Daten komplexer genetischer Erkrankungen zu verwenden, ist lediglich in DE-19543065 A1 erwähnt. So wird hierin zum Beispiel die eigentliche Detektion nicht über die Analyse der Hybrisierung von Nuklein-säure-Sonden im Massenspektrometer durchgeführt.

35

Nicht immer ist es erforderlich, tatsächlich die gesamte Sequenz eines Gens oder Genabschnitts zu ermitteln, wie dies bei einer Sequenzierung das Ziel ist. Dies ist insbesondere dann der Fall, wenn wenige 5-Methylcytosin-  
5 Positionen innerhalb einer längeren Basensequenz für eine Vielzahl von unterschiedlichen Proben abzutasten sind. Die Sequenzierung liefert hier in großem Umfang redundante Information und ist zudem sehr teuer. Dies ist auch schon dann der Fall, wenn die Sequenz bereits bekannt ist  
10 und ausschließlich Methylierungspositionen dargestellt werden sollen. Auch ist es denkbar, daß in einigen Fällen überhaupt nur die Unterschiede im Methylierungsmuster zwischen verschiedenen genomischen DNA-Proben von Interesse sind und daß auf die Ermittlung einer Vielzahl über-einstimmender methylierter Positionen verzichtet werden kann. Für die hier angeführten Fragestellungen existiert bislang kein Verfahren, das ohne Sequenzierung jeder einzelnen Probe, kostengünstig die gewünschten Ergebnisse liefert.

20 Sequenzinformation muß auch umso mehr immer weniger neu ermittelt werden, da die Genomprojekte, deren Ziel die vollständige Sequenz verschiedener Organismen ist, zügig voranschreiten. Vom menschlichen Genom sind zwar derzeit erst etwa 5 % fertig sequenziert, jedoch kommen jetzt,  
25 weil andere Genomprojekte dem Ende zuneigen und dadurch Sequenzierressourcen frei werden, jedes Jahr weitere 5 % dazu. Mit der Vervollständigung der Sequenzierung des menschlichen Genoms wird bis zum Jahre 2006 gerechnet.

30 Matrix-assistierte Laser Desorptions/Ionisations Massenspektrometrie (MALDI) ist eine neue, sehr leistungsfähige Entwicklung für die Analyse von Biomolekülen (Karas, M. and Hillenkamp, F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10.000 daltons. Anal. Chem. 60: 2299-2301). Ein Analytmolekül wird in ei-

ne im UV absorbierende Matrix eingebettet. Durch einen kurzen Laserpuls wird die Matrix ins Vakuum verdampft und <sup>5</sup> das Analyt so unfragmentiert in die Gasphase befördert. Eine angelegte Spannung beschleunigt die Ionen in ein feldfreies Flugrohr. Auf Grund ihrer verschiedenen Massen werden Ionen unterschiedlich stark beschleunigt. Kleinere Ionen erreichen den Detektor früher als größere. Die Flugzeit wird in die Masse der Ionen umgerechnet.

10 Technische Neuerungen der Hardware haben die Methode signifikant verbessert. Erwähnenswert ist hierbei die „delayed extraction“-Methode (DE). Für DE wird die Beschleunigungsspannung mit einer Verzögerung zum Laserpuls eingeschaltet und dadurch eine verbesserte Auflösung der Signale erreicht, weil die Zahl der Stöße verringert <sup>15</sup> wird.

MALDI eignet sich ausgezeichnet zur Analyse von Peptiden und Proteinen. Für Nukleinsäuren ist die Empfindlichkeit <sup>20</sup> etwa 100-mal schlechter als für Peptide und nimmt mit zunehmender Fragmentgröße überproportional ab. Der Grund dafür liegt darin, daß für die Ionisation von Peptiden und Proteinen lediglich ein einziges Proton eingefangen werden muß. Für Nukleinsäuren, die ein vielfach negativ <sup>25</sup> geladenes Rückgrat haben, ist der Ionisationsprozeß durch die Matrix wesentlich ineffizienter. Für MALDI spielt die Wahl der Matrix eine eminent wichtige Rolle. Für die Desorption von Peptiden sind einige sehr leistungsfähige Matrices gefunden worden, die eine sehr feine Kristallisation ergeben. Für DNA sind zwar inzwischen einige geeignete Matrices gefunden worden, jedoch wurde dadurch <sup>30</sup> der Empfindlichkeitsunterschied nicht verringert. Phosphorothioatnukleinsäuren, bei denen die gewöhnlichen Phosphate des Rückgrats durch Thiophosphate substituiert sind, lassen sich durch einfache Alkylierungsschemie in <sup>35</sup> eine ladungsneutrale DNA umwandeln.

Die Kopplung eines "charge tags" an diese modifizierte DNA resultiert in der Steigerung der Empfindlichkeit in den gleichen Bereich wie er für Peptide gefunden wird.

5 Durch diese Modifikationen hat sich auch die Möglichkeit eröffnet, ähnliche Matrices, wie sie für die Desorption von Peptiden verwendet werden, zu nutzen. Ein weiterer Vorteil von charge tagging ist die erhöhte Stabilität der Analyse gegen Verunreinigungen, die den Nachweis unmodifizierter Substrate stark erschweren. PNAs und Methylphosphonatoligonukleotide sind mit MALDI untersucht worden und lassen sich so analysieren.

15 Derzeit ist diese Technologie in der Lage, im Massenbereich von 1.000 bis 4.000 Da, Moleküle mit einer Massendifferenz von 1 Da zu unterscheiden. Durch die natürliche Verteilung von Isotopen sind die meisten Biomoleküle jedoch schon etwa 5 Da breit. Technisch ist diese massenspektrometrische Methode also vorzüglich für die Analyse 20 von Biomolekülen geeignet. Vernünftigerweise müssen zu analysierende Produkte, die unterschieden werden sollen, also mindestens 5 Da auseinander liegen. In diesem Massenbereich könnten also 600 Moleküle unterschieden werden.

25 Ein Array mit vielen tausend Ziel-DNAs kann auf einer Festphase immobilisiert und anschließend können alle Ziel-DNAs gemeinsam auf das Vorhandensein einer Sequenz mittels einer Sonde (Nukleinsäure mit komplementärer Sequenz) untersucht werden.

30 Eine Übereinstimmung in der Ziel-DNA mit der Sonde kommt durch eine Hybridisation der zwei Teile miteinander zu-stande. Sonden können beliebige Nukleinsäuresequenzen von beliebiger Länge sein. Es existieren verschiedene Verfah-

ren für die Auswahl von optimalen Bibliotheken von Son-  
densequenzen, welche minimal miteinander überlappen.  
Sondensequenzen können auch gezielt zusammengestellt wer-  
den, um bestimmte Ziel-DNA Sequenzen aufzufinden. Oligo-  
fingerprinting ist ein Ansatz, bei welchem diese Techno-  
logie zum Einsatz kommt. Eine Bibliothek von Ziel-DNAs  
wird mit kurzen Nukleinsäuresonden abgetastet. Meist sind  
hier die Sonden nur 8-12 Basen lang. Es wird jeweils eine  
Sonde auf einmal an eine auf einer Nylonmembran immobili-  
sierte Ziel-DNA-Bibliothek hybridisiert. Die Sonde ist  
radioaktiv markiert und die Hybridisierung wird anhand  
der Lokalisierung der Radioaktivität beurteilt. Für die  
Abtastung eines immobilisierten DNA-Arrays sind auch  
schon fluoreszent markierte Sonden verwendet worden.

US 5,605,798 beschreibt die Abtastung von über Hybridisierung immobilisierten Ziel-Nukleinsäuren mit Nuklein-  
säuresonden und Massenspektrometrie. Hierbei wird jedoch  
weder gezielt eine Identifikation von Methylierungsmu-  
stern durchgeführt, noch werden modifizierte Nukleinsäu-  
ren (z. B. PNAs, Charge Tags) eingesetzt, noch wird eine  
Genomamplifikation durchgeführt.

Als Sonden kommen jegliche Moleküle in Frage, die se-  
quenzspezifisch mit einer Ziel-DNA wechselwirken können.  
Am gängigsten sind Oligodeoxyribonukleotide. Jedoch bie-  
tet sich jede Modifikation von Nukleinsäuren an, z.B.  
Peptide Nucleic Acids (PNA), Phosphorothioatoligonukleo-  
tide oder Methylphosphonatoligonukleotide. Die Spezifität  
einer Sonde ist sehr wesentlich. Phosphorothioatoligonu-  
kleotide sind nicht allzugut geeignet, da ihre Struktur  
durch die Schwefelatome und dadurch die Hybridisie-  
rungseigenschaft gestört ist. Ein Grund hierfür könnte  
sein, daß Phosphorothioatoligonukleotide normalerweise  
nicht diastereomerenrein synthetisiert werden. Bei Me-  
thylphosphonatoligonukleotiden besteht ein ähnliches Pro-

blem, jedoch werden diese Oligonukleotide vermehrt dia-  
stereomerenrein synthetisiert. Ein wesentlicher Unter-  
schied dieser Modifikation ist das ungeladene Rückgrat,  
welches zu einer verminderten Abhängigkeit der Hybridisa-  
tion von Puffersalzen und insgesamt durch die geringere  
Abstoßung zu höherer Affinität führt. Peptide Nucleic  
Acids haben ebenfalls ein ungeladenes Rückgrat, welches  
gleichzeitig chemisch sehr stark von der gängigen Zucker-  
Phosphat Struktur des Rückgrats in Nukleinsäuren ab-  
weicht. Das Rückgrat einer PNA hat eine Amidsequenz an-  
stelle des Zucker-Phosphat Rückgrats gewöhnlicher DNA.  
PNA hybridisiert sehr gut mit DNA komplementärer Sequenz.  
Die Schmelztemperatur eines PNA/DNA-Hybrids ist höher als  
die des entsprechenden DNA/DNA-Hybrids und wiederum ist  
die Abhängigkeit der Hybridisierung von Puffersalzen re-  
lativ gering.

Kombinatorische Synthesen, d.h. die Herstellung von Sub-  
stanzbibliotheken ausgehend von einem Gemisch von Vorstu-  
fen, werden sowohl auf fester als auch in flüssiger Phase  
durchgeführt. Vor allem die kombinatorische Festphasen-  
synthese hat sich frühzeitig etabliert, da in diesem Fall  
die Abtrennung von Nebenprodukten besonders einfach ist.  
Nur die an den Support gebundenen Zielverbindungen werden  
in einem Waschschritt zurück behalten und am Ende der Syn-  
these durch das gezielte Spalten eines Linkers isoliert.  
Diese Technik erlaubt auf einfachem Wege die gleichzeiti-  
ge Synthese einer Vielzahl verschiedener Verbindungen an  
einer Festphase und somit den Erhalt von chemisch  
30 "reinen" Substanzbibliotheken.

Daher sind die Verbindungsklassen, die auch in nicht kom-  
binatorischen, konventionellen Synthesen auf einer Fest-  
phase synthetisiert werden, der kombinatorischen Chemie  
35 besonders leicht zugänglich und werden demzufolge auch

breit verwendet. Dies trifft vor allem auf Peptid-, Nukleinsäure- und PNA-Bibliotheken zu.

Die Synthese von Peptiden erfolgt durch Binden der ersten  
5 N-geschützten Aminosäure (z.B. Boc) an den Support, nach-  
folgende Entschützung und Reaktion der zweiten Aminosäure  
mit der freigewordenen NH<sub>2</sub>-Gruppe der ersten. Nicht rea-  
gierte Aminofunktionen werden in einem weiteren "Capping"  
10 Schritt einer Weiterreaktion im nächsten Synthesezyklus  
entzogen. Die Schutzgruppe an der Aminofunktion der zwei-  
ten Aminosäure wird entfernt und der nächste Baustein  
kann gekoppelt werden. Zur Synthese von Peptidbiblio-  
theken wird ein Gemisch von Aminosäuren in einem oder mehre-  
ren Schritten verwendet. Die Synthese von PNA und PNA-  
15 Bibliotheken erfolgt sinngemäß.

Nukleinsäure-Bibliotheken werden meist durch Festphasen-  
synthese mit Gemischen verschiedener Phosphoramidit-  
Nukleoside erhalten. Dies kann auf kommerziell erhältli-  
20 chen DNA-Synthesizern ohne Veränderungen in den Synthese-  
protokollen durchgeführt werden.

Verschiedene Arbeiten zur kombinatorischen Synthese von  
PNA Bibliotheken sind publiziert worden. Diese Arbeiten  
25 behandeln den Aufbau von kombinatorischen Sequenzen, d.h.  
die Synthese von PNAs in denen einzelne, spezifische Ba-  
sen in der Sequenz durch degenerierte Basen ersetzt wer-  
den und dadurch zufällige Sequenzvarianz erreicht wird.

30 Die Verwendung massenspektrometrischer Methoden für die  
Analyse kombinatorischer Bibliotheken ist mehrfach be-  
schrieben worden.

35 Es existieren verschiedene Verfahren um DNA zu immobili-  
sieren. Das bekannteste Verfahren ist die Festbindung ei-  
ner DNA, welche mit Biotin funktionalisiert ist, an eine

Streptavidin-beschichtete Oberfläche. Die Bindungsstärke dieses Systems entspricht einer kovalenten chemischen Bindung ohne eine zu sein. Um eine Ziel-DNA kovalent an eine chemisch vorbereitete Oberfläche binden zu können, bedarf es einer entsprechenden Funktionalität der Ziel-DNA. DNA selbst besitzt keine Funktionalisierung, die geeignet ist. Es gibt verschiedene Varianten in eine Ziel-DNA eine geeignete Funktionalisierung einzuführen: Zwei leicht zu handhabende Funktionalisierungen sind primäre, aliphatische Amine und Thiole. Solche Amine werden quantitativ mit N-Hydroxy-succinimidestern umgesetzt und Thiole reagieren unter geeigneten Bedingungen quantitativ mit Alkyliodiden. Eine Schwierigkeit besteht im Einführen einer solchen Funktionalisierung in eine DNA. Die einfachste Variante ist die Einführung durch einen Primer einer PCR. Gezeigte Varianten benutzen 5'-modifizierte Primer (NH<sub>2</sub> und SH) und einen bifunktionalen Linker.

Ein wesentlicher Bestandteil der Immobilisierung auf einer Oberfläche ist ihre Beschaffenheit. Bis jetzt beschriebene Systeme sind hauptsächlich aus Silizium oder Metall (magnetic beads). Eine weitere Methode zur Bindung einer Ziel-DNA basiert darauf, eine kurze Erkennungssequenz (z.B. 20 Basen) in der Ziel-DNA zur Hybridisierung an ein oberflächenimmobilisiertes Oligonukleotid zu verwenden.

Es sind auch enzymatische Varianten zur Einführung von chemisch aktivierten Positionen in eine Ziel-DNA beschrieben worden. Hier wird an einer Ziel-DNA enzymatisch eine 5'-NH<sub>2</sub>-Funktionalisierung durchgeführt.

Aufgabe der vorliegenden Erfindung ist es, ein Verfahren zu schaffen, welche die Nachteile des Standes der Technik überwindet und in der Lage ist, effektiv und hochparalle-

lisiert Cytosin-Methylierungen in einem Array von immobilisierten genomischen DNA-Proben aufzuzeigen.

Gegenstand der vorliegenden Erfindung ist daher ein Verfahren zur Auffindung von epigenetischen Informationsträgern in Form von 5-Methylcytosin-Basen in genomischer DNA, das eine Vielzahl von Sonden gleichzeitig zur massenspektrometrischen Untersuchung eines Arrays von Ziel-Nukleinsäuren verwendet.

10

Die erfindungsgemäße Aufgabe wird gelöst, in dem ein Verfahren Verfahren zur Identifikation von Cytosin-Methylierungsmustern in genomischen DNA-Proben zur Verfügung gestellt wird, bei dem man:

15

a) eine genomische DNA-Probe chemisch derart behandelt, daß Cytosin und 5-Methylcytosin unterschiedlich reagieren und man in der Duplex ein unterschiedliches Basenpaarungsverhalten der beiden Produkte erhält;

20

b) Teile der so behandelten DNA-Probe enzymatisch amplifiziert;

25

c) die amplifizierten Teile der so behandelten DNA-Probe an eine Oberfläche bindet;

30

d) einen Satz von Sonden unterschiedlicher Nukleobasensequenzen, die jeweils mindestens einmal die Dinukleotidsequenz 5'-CpG-3' enthalten, an die immobilisierten DNA-Proben hybridisiert;

e) die nicht hybridisierten Sonden abtrennt;

35

f) die hybridisierten Sonden in einem Massenspektrometer analysiert, wobei die Position der Sonden auf dem Probenträger eine Zuordnung zu der hybridisierenden DNA-Probe

erlaubt;

5 g) Übertragung der aus den Massenspektren erhaltenen Pe-  
akmuster in Methylierungsmuster und Abgleich der neuen  
Daten mit einer Datenbank.

Erfnungsgemäß bevorzugt ist es, daß man in c) ein oder  
mehrere amplifizierte genomische DNA-Fragmente durch Hy-  
bridisierung mit komplementären Oligonukleotid- oder PNA-  
10 Sequenzen, die kovalent an die Oberfläche gebunden sind,  
immobilisiert.

Erfnungsgemäß bevorzugt ist es ferner, daß nach der Hy-  
bridisierung ein Cross-Linking der genomischen DNA-  
15 Fragmente mit den an der Oberfläche gebundenen Oligonu-  
kleotid- oder PNA-Sequenzen erfolgt. Besonders bevorzugt  
ist hierbei, daß man zum Cross-Linking kovalente chemi-  
sche Bindungen ausbildet. Erfnungsgemäß bevorzugt ist  
hierbei ferner, daß man zum Cross-Linking elektrostati-  
20 sche Wechselwirkungen ausbildet.

Vorteilhaft ist es ferner, daß die an die Oberfläche ge-  
bundenen Oligonukleotid- oder PNA-Sequenzen 5-Bromuracil-  
Bausteine enthalten.

25 Erfnungsgemäß bevorzugt ist es, daß die immobilisierten  
komplementären Oligonukleotidsequenzen modifizierte Ba-  
sen, Ribose oder Rückgrateinheiten beinhalten.

30 Das erfungsgemäße Verfahren ist ferner dadurch gekenn-  
zeichnet, daß man in b) die genomische DNA-Probe in Form  
mehrerer amplifizierter Fragmente vermehrt, so daß man  
mindestens 0,01 % des gesamten Genoms amplifiziert.

35 Erfnungsgemäß bevorzugt ist es auch, daß man die Mi-  
schung amplifizierter DNA-Fragmente auf eine Oberfläche

bindet, auf der eine Vielzahl von unterschiedlichen Punkten angeordnet ist die jeweils in der Lage sind, unterschiedliche Teile der amplifizierten DNA-Probe zu binden.

5 Erfnungsgemäß bevorzugt ist ferner, daß man in d) einen Satz von Sonden verwendet, welcher die Dinukleotid-Sequenz 5'-CpG-3' nur einmal je Sonde enthält und die Sonden ansonsten jeweils entweder keine Cytosin oder keine Guanin Basen umfassen.

10

Außerdem ist es erfungsgemäß bevorzugt, daß man Schritt a) eine Bisulfit- oder Pyrosulfit- oder Disulfit-lösung oder eine Mischung aus den angegebenen Lösungen, zusammen mit anderen Reagenzien für die spezifische oder 15 hinreichend selektive Umwandlung von Cytosin in Uracil verwendet.

Vorteilhaft ist es auch, daß die zur Immobilisierung von 20 amplifizierter Proben-DNA verwendete Oberfläche zugleich der Probenträger für ein Massenspektrometer ist. Bevorzugt ist es dabei, daß man die zur Immobilisierung von amplifizierter Proben-DNA verwendete Oberfläche als ganzes vor f) auf einen Probenträger für ein Massenspektrometer aufbringt. Bevorzugt ist es hierbei auch, daß man 25 die hybridisierten Sonden vor dem, nach dem oder durch den Kontakt mit einer Matrix von den immobilisierten amplifizierten DNA-Proben ablöst.

Erfnungsgemäß bevorzugt ist es ferner, daß die Sonden Nukleinsäuren sind, die ein oder mehrere Massen-Tags tragen. Erfnungsgemäß vorteilhaft ist es auch, daß ein oder mehrere Massen-Tags zugleich Ladungs-Tags sind. Oder daß die Sonden zusätzlich ein Ladungs-Tag tragen.

35 Erfnungsgemäß bevorzugt ist es, daß die Sonden modifizierte Nukleinsäuremoleküle sind. Hierbei ist besonders

bevorzugt, daß die modifizierten Nukleinsäuremoleküle PNAs, alkylierte Phosphorothioatnukleinsäuren oder Alkylphosphonatnukleinsäuren sind.

5 Erfindungsgemäß bevorzugt ist es, daß man die Sonden durch kombinatorische Synthese herstellt. Dabei ist erfundungsgemäß besonders bevorzugt, daß man unterschiedliche Basenbausteine derart markiert, daß die aus ihnen synthetisierten Sonden jeweils über ihre Massen im Massenspektrometer unterscheidbar sind.

10 15 Erfindungsgemäß vorteilhaft ist es ferner, daß man die Sonden als Teilbibliotheken herstellt und diese mit unterschiedlichen Massen- und/oder Ladungs-Tags versieht.

Ganz besonders erfundungsgemäß bevorzugt ist es, daß man in f) matrix-assistierte Laser Desorptions/Ionisations Massenspektrometrie (MALDI) ausführt.

20 25 Ein weiterer Gegenstand der vorliegenden Erfindung ist ein Kit für die Durchführung des erfundungsgemäßen Verfahrens, umfassend einen Probenträger für ein Massenspektrometer, welcher derart modifiziert ist, daß beliebig wählbare Teile eines Genoms an diesen immobilisiert werden und/oder Sondenbibliotheken, mit denen man die an den Probenträger immobilisierte DNA massenspektrometrisch analysiert und/oder weiteren Chemikalien, Lösungsmitteln und/oder Hilfsstoffen sowie gegebenenfalls einer Bedienungsanleitung.

30 35 Das erfundungsgemäße Verfahren dient zur Identifikation von 5-Methylcytosin-Positionen in genomischer DNA, die unterschiedlichsten Ursprungs sein kann. Die genomische DNA wird zunächst chemisch so behandelt, daß sich ein Unterschied in der Reaktion der Cytosin-Basen zu den 5-Methylcytosin-Basen ergibt. Mögliche Reagenzien sind hier

z. B. Disulfit (auch als Bisulfit bezeichnet), Hydrazin und Permanganat. In einer bevorzugten Variante des Verfahrens wird die genomische DNA mit Disulfit in Gegenwart von Hydrochinon oder Hydrochinonderivaten behandelt, wobei selektiv nach anschließender alkalischer Hydrolyse die Cytosin-Basen in Uracil umgewandelt werden. 5-Methylcytosin bleibt unter diesen Bedingungen unverändert. Nach einem Aufreinigungsprozeß, der der Abtrennung des überschüssigen Disulfits dient, werden bestimmte Abschnitte der vorbehandelten genomischen DNA in einer Polymerasereaktion amplifiziert. In einer bevorzugten Variante des Verfahrens wird hier die Polymerasekettenreaktion eingesetzt. In einer besonders bevorzugten Variante des Verfahrens wird die Polymerasekettenreaktion so durchgeführt, daß mindestens 0.01% des gesamten Genoms in Form mehrerer Fragmente amplifiziert werden.

Die amplifizierte, vorbehandelte DNA-Probe kann nun in mehreren Varianten des Verfahrens an eine Oberfläche immobilisiert werden. In einer bevorzugten Variante des Verfahrens erfolgt die Immobilisierung an die Oberfläche derart, daß diese zuvor mit Oligonukleotiden oder kurzen PNA (Peptide nucleic acids) -Sequenzen modifiziert wurde und damit eine Hybridisierung komplementärer Sequenzen in der DNA-Probe erfolgt. Grundsätzlich können die immobilisierten Oligonukleotide sowohl an den Basen, an der (Desoxy)-Ribose und/oder am Rückgrat gegenüber herkömmlicher DNA modifiziert sein. Werden nun verschiedene Oligonukleotid- oder PNA-Sequenzen in Form eines Arrays an diese Oberfläche gebunden oder auf ihr synthetisiert, so kann jede dieser unterschiedlichen Sequenzen unterschiedliche Teile der amplifizierten DNA-Fragmente binden. In einer besonders bevorzugten Variante des Verfahrens wird im Anschluß an die Hybridisierung ein Cross-Linking der beiden Stränge durchgeführt. Dies kann durch Ausbildung einer kovalenten chemischen Bindung oder einer stabilen

elektrostatischen Wechselwirkung erfolgen. In einer weiteren bevorzugten Variante wird ein photochemisches Cross-Linking über Bromuracil-Bausteine durchgeführt.

5 Auch ist es möglich, Fragmente der vorbehandelten genomischen DNA separat zu amplifizieren und die Produkte einzeln an unterschiedlichen Orten auf der Oberfläche zu immobilisieren. In einer bevorzugten Variante des Verfahrens erfolgt dies derart, daß einer der PCR-Primer eine  
10 zur Immobilisierung geeignete Funktion trägt, die mit einer auf der Oberfläche angebrachten Funktionalität eine Bindung eingehen kann.

Die Oberfläche, an die die amplifizierten DNA-Fragmente gebunden werden, soll sich entweder auf den Probenträger eines MALDI-Massenspektrometers übertragen lassen oder aber selbst dieser Probenträger sein. Die Konstruktion und Software des Massenspektrometers gewährleistet dabei, daß der jeweils untersuchte Punkt auf dem Probenträger  
20 den dort ursprünglich gebundenen Proben zugeordnet werden kann.

An die immobilisierten amplifizierten DNA-Fragmente wird nun ein Satz von Sonden hybridisiert, wobei diese Sonden je wenigstens einmal die Sequenz 5'-CpG-3' und ansonsten jeweils entweder keine Cytosin oder Guanin Basen enthalten. Die Sonden können Oligonukleotide, modifizierte Oligonukleotide oder PNAs (Peptide nucleic acids) sein. In einer bevorzugten Variante des Verfahrens wird dieser Satz von Sonden in einem kombinatorischen Syntheseansatz als kombinatorische Bibliothek hergestellt. In einer weiteren bevorzugten Variante des Verfahrens sind die Sonden über ihre Massen eindeutig unterscheidbar, so daß ein Rückschluß von der Masse auf die Sequenz möglich ist. Dazu  
30 können die Sonden mit Massentags versehen sein, die die  
35 Massengleichheit unterschiedlicher Sonden verhindern. Die

Sonden können mit Ladungstags versehen sein, um eine bessere Darstellbarkeit im Massenspektrometer zu erreichen und um die Stabilität der Analyse in Anwesenheit von Salzen und Detergenzien zu erhöhen. Die Massentags können zugleich Ladungstags sein. Die Sonden können auch als kombinatorische Teilbiblioteken hergestellt werden, die wiederum unterschiedliche Massen- und/oder Ladungstags tragen. Die Sonden können PNAs, unmodifizierte Nukleinsäuremoleküle oder modifizierte Nukleinsäuremoleküle wie Phosphothioatnukleinsäuren, alkylierte Phosphorothioatnukleinsäuren oder Alkylphosphonatnukleinsäuren sein, unbeschadet einer weiteren Modifikation durch Massen- und Ladungstags.

Die nicht hybridisierten Sonden werden in einem oder mehreren Waschschriften abgetrennt. Die hybridisierten Sonden bleiben dabei an ihren Positionen.

Die Oberfläche wird auf dem MALDI-Probenträger befestigt und in das Massenspektrometer transferiert bzw. unmittelbar transferiert, wenn das Verfahren auf dem MALDI-Probenträger selbst durchgeführt wurde. Der Array von Proben wird nun auf hybridisierte Sonden hin massenspektrometrisch untersucht. Die hybridisierten Sonden werden dazu in einer bevorzugten Variante des Verfahren durch den Kontakt mit der MALDI-Matrix dehybridisiert und in diese eingebettet; durch die Geschwindigkeit in der die Matrix aufgebracht wird erfolgt aber keine gegenseitige Kontamination der benachbarten Punkte. An jedem Punkt ergeben die hybridisierten Sonden ein Peakmuster, über das auf die Sequenzen geschlossen werden kann, an denen eine Hybridisierung erfolgte. Aufgrund der Vorbehandlung (vorzugsweise mit Bisulfat) ergeben unterschiedlich am Cytosin methylierte DNA-Fragmente unterschiedliche Sequenzen. Daher sind die durch die Sonden erzeugten Peakmuster im Massenspektrometer charakteristische Methylier-

rungsmuster der jeweils untersuchten DNA-Probe. Es erfolgt ein Abgleich dieser Methylierungsmuster mit denen einer Datenbank.

## Patentansprüche

1. Verfahren zur Identifikation von Cytosin-Methylierungsmustern in genomischen DNA-Proben dadurch gekennzeichnet, daß man:
  - a) eine genomische DNA-Probe chemisch derart behandelt, daß Cytosin und 5-Methylcytosin unterschiedlich reagieren und man in der Duplex ein unterschiedliches Basenpaarungsverhalten der beiden Produkte erhält;
  - b) Teile der so behandelten DNA-Probe enzymatisch amplifiziert;
  - c) die amplifizierten Teile der so behandelten DNA-Probe an eine Oberfläche bindet;
  - d) einen Satz von Sonden unterschiedlicher Nukleobase-Sequenzen, die jeweils mindestens einmal die Dinukleotidsequenz 5'-CpG-3' enthalten, an die immobilisierten DNA-Proben hybridisiert;
  - e) die nicht hybridisierten Sonden abtrennt;
  - f) die hybridisierten Sonden in einem Massenspektrometer analysiert, wobei die Position der Sonden auf dem Probenträger eine Zuordnung zu der hybridisierten DNA-Probe erlaubt;
  - g) Übertragung der aus den Massenspektren erhaltenen Peakmuster in Methylierungsmuster und Abgleich der neuen Daten mit einer Datenbank.
- 35 2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß man in c) ein oder mehrere amplifizierte genomi-

sche DNA-Fragmente durch Hybridisierung mit komplementären Oligonukleotid- oder PNA-Sequenzen, die kovalent an die Oberfläche gebunden sind, immobilisiert.

5

3. Verfahren nach Anspruch 2, dadurch gekennzeichnet, daß nach der Hybridisierung ein Cross-Linking der genomischen DNA-Fragmente mit den an der Oberfläche gebundenen Oligonukleotid- oder PNA-Sequenzen erfolgt.
4. Verfahren nach Anspruch 3, dadurch gekennzeichnet, daß man zum Cross-Linking kovalente chemische Bindungen ausbildet.
- 15 5. Verfahren nach Anspruch 3, dadurch gekennzeichnet, daß man zum Cross-Linking elektrostatische Wechselwirkungen ausbildet.
6. Verfahren nach einem der Ansprüche 3 bis 5, dadurch gekennzeichnet, daß die an die Oberfläche gebundenen Oligonukleotid- oder PNA-Sequenzen 5-Bromuracil-Bausteine enthalten.
- 20 7. Verfahren nach mindestens einem der voranstehenden Ansprüche, dadurch gekennzeichnet, daß die immobilisierten komplementären Oligonukleotidsequenzen modifizierte Basen, Ribose oder Rückgrateinheiten beinhalten.
- 25 8. Verfahren nach einem der voranstehenden Ansprüche, dadurch gekennzeichnet, daß man in b) die genomische DNA-Probe in Form mehrerer amplifizierter Fragmente vermehrt, so daß man mindestens 0,01 % des gesamten Genoms amplifiziert.

9. Verfahren nach mindestens einem der voranstehenden Ansprüche, dadurch gekennzeichnet, daß man die Mischung amplifizierter DNA-Fragmente auf eine Oberfläche bindet, auf der eine Vielzahl von unterschiedlichen Punkten angeordnet ist die jeweils in der Lage sind, unterschiedliche Teile der amplifizierten DNA-Probe zu binden.  
5
10. Verfahren nach einem der voranstehenden Ansprüche, dadurch gekennzeichnet, daß man in d) einen Satz von Sonden verwendet, welcher die Dinukleotid-Sequenz 5'-CpG-3' nur einmal je Sonde enthält und die Sonden ansonsten jeweils entweder keine Cytosin oder keine Guanin Basen enthalten.  
15
11. Verfahren nach einem der voranstehenden Ansprüche, dadurch gekennzeichnet, daß man Schritt a) eine Bisulfit- oder Pyrosulfit- oder Disulfitlösung oder eine Mischung aus den angegebenen Lösungen, zusammen mit anderen Reagenzien für die spezifische oder hinreichend selektive Umwandlung von Cytosin in Uracil verwendet.  
20
12. Verfahren nach einem der voranstehenden Ansprüche, dadurch gekennzeichnet, daß die zur Immobilisierung von amplifizierter Proben-DNA verwendete Oberfläche zugleich der Probenträger für ein Massenspektrometer ist.  
25
- 30 13. Verfahren nach mindestens einem der Ansprüche 1 bis 11, dadurch gekennzeichnet, daß man die zur Immobilisierung von amplifizierter Proben-DNA verwendete Oberfläche als ganzes vor f) auf einen Probenträger für ein Massenspektrometer aufbringt.

14. Verfahren nach einem der Ansprüche 1 bis 13, dadurch gekennzeichnet, daß man die hybridisierten Sonden vor dem, nach dem oder durch den Kontakt mit einer Matrix von den immobilisierten amplifizierten DNA-Proben ab-  
5 löst.
15. Verfahren nach einem der voranstehenden Ansprüche, dadurch gekennzeichnet, daß die Sonden Nukleinsäuren sind, die ein oder mehrere Massen-Tags tragen.  
10
16. Verfahren nach Anspruch 15, dadurch gekennzeichnet, daß ein oder mehrere Massen-Tags zugleich Ladungs-  
Tags sind.
- 15 17. Verfahren nach Anspruch 15, dadurch gekennzeichnet, daß die Sonden zusätzlich ein Ladungs-Tag tragen.
18. Verfahren nach einem der voranstehenden Ansprüche, dadurch gekennzeichnet, daß die Sonden modifizierte  
20 Nukleinsäuremoleküle sind.
19. Verfahren nach Anspruch 20, dadurch gekennzeichnet, daß die modifizierten Nukleinsäuremoleküle PNAs, al-  
kylierte Phosphorothioatnukleinsäuren oder Alkylphos-  
25 phonatnukleinsäuren sind.
20. Verfahren nach einem der voranstehenden Ansprüche, dadurch gekennzeichnet, daß man die Sonden durch kom-  
binatorische Synthese herstellt.  
30
21. Verfahren nach Anspruch 20, dadurch gekennzeichnet, daß man unterschiedliche Basenbausteine derart mar-  
kiert, daß die aus ihnen synthetisierten Sonden je-  
weils über ihre Massen im Massenspektrometer unter-  
35 scheidbar sind.

22. Verfahren nach einem der voranstehenden Ansprüche, dadurch gekennzeichnet, daß man die Sonden als Teilbibliotheken herstellt und diese mit unterschiedlichen Massen- und/oder Ladungs-Tags versieht.

5

23. Verfahren nach einem der voranstehenden Ansprüche, dadurch gekennzeichnet, daß man in f) matrix-assistierte Laser Desorptions/Ionisations Massenspektrometrie (MALDI) ausführt.

10

24. Kit für die Durchführung des Verfahrens nach Anspruch 1, umfassend einen Probenträger für ein Massenspektrometer, welcher derart modifiziert ist, daß beliebig wählbare Teile eines Genoms an diesen immobilisiert werden und/oder Sondenbibliotheken, mit denen man die an den Probenträger immobilisierte DNA massenspektrometrisch analysiert und/oder weiteren Chemikalien, Lösungsmitteln und/oder Hilfsstoffen sowie gegebenenfalls einer Bedienungsanleitung.

15

20